**NumPy (Numerical Python)**

(arrays and functions on arrays)

NumPy is a library for python language which adds support for large, multi-dimensional arrays and matrices, along with large collection of high-level mathematical functions to operate on these arrays.

**Example:**

data.toarray() converts the data to an array

**Pandas**

Pandas is a library for python language mainly for data manipulation and analysis.

It provides data structures and operations for manipulating numerical tables and datasets.

**Example:**

Df = pd.DataFrame(data) create a dataframe from data(array)

movies = pd.read\_csv(‘location/dataset’) read dataset as a dataframe

dataframe.head() shows first 5 rows of dataframe

dataframe.head(1) shows first row

df1.merge(df2, on=’column’) merge 2 dataframes based on specified column

dataframe.shape shows shape of dataframe (rows x cols)

dataframe.iloc[0] list 0th row from the dataframe

**AST (Abstract syntax tree)**

**Example:**

ast.literal\_eval(string object) It converts string object to list

**Scikit-learn**

Sklearn is a machine learning library which provides support for various machine learning tools like vector, etc.

**Example:**

countVectorizer class get vector for words occurring in each tag out of 5000 words (0=no, 1=yes)

cv = countVectorizer(max\_features=5000, stop\_words=’english’)

cv.fit\_transform(tags column).toarray()

cosine\_similarity class get distance of each vector with every other vector

cosine\_similarity(vectors)

**NLTK (Natural Language Toolkit)**

Python library to process English language.

**Example:**

PorterStemmer class remove similar words, convert them to standard (love, loving, loved=love)

ps = PorterStemmer()

**Pickle**

Python pickle module is used for serializing and de-serializing python object structures.

pickle.dump(dataset.to\_dict(), open('dataset.pkl','wb')) send dataset to folder as dictionary